

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Application of Signal Processing in Power Quality Monitoring

Zahra Moravej<sup>1</sup>, Mohammad Pazoki<sup>2</sup> and Ali Akbar Abdoos<sup>3</sup>

<sup>1,2</sup>*Electric and computer engineering faculty, Semnan University,*

<sup>3</sup>*Electric and computer engineering faculty, Babol Noshirvani University of Technology  
Iran*

## 1. Introduction

The definition of power quality according to the Institute of Electrical and Electronics Engineers (IEEE) dictionary [159, page 807] is that “power quality is the concept of powering and grounding sensitive equipment in a matter that is suitable to the operation of that equipment.”

In recent years, power quality (PQ) has become a significant issue for both utilities and customers. PQ issues and the resulting problems are the consequences of the increasing use of solid-state switching devices, non-linear and power electronically switched loads, unbalanced power systems, lighting controls, computer and data processing equipment, as well as industrial plant rectifiers and inverters. These electronic-type loads cause quasi-static harmonic dynamic voltage distortions, inrush, pulse-type current phenomenon with excessive harmonics, and high distortion. A PQ problem usually involves a variation in the electric service voltage or current, such as voltage dips and fluctuations, momentary interruptions, harmonics, and oscillatory transients, causing failure or inoperability of the power service equipment. Hence, to improve PQ, a fast and reliable detection of disturbances and sources and causes of such disturbances must be known before any appropriate mitigating action can be taken.

The algorithms for detection and classification of power quality disturbances (PQDs) are generally divided into three main steps: (1) generation of PQDs, (2) feature extraction, and (3) classification of extracted vectors (Uyara et al., 2008; Gaing 2004; Moravej et al., 2010; Moravej et al., 2011a).

It will be evident that the electricity supply waveform, often thought of as composed of pure sinusoidal quantities, can suffer a wide variety of disturbances. Mathematical equations and simulation software such as MATLAB simulink (MATLAB), PSCAD/EMTDC (PSCAD/EMTDC 1997), ATP (ATPDraw for Windows 1998) are usually used for generation of PQ events. To make the study of Power Quality problems useful, the various types of disturbances need to be classified by magnitude and duration. This is especially important for manufacturers and users of equipment that may be at risk. Manufacturers need to know what is expected of their equipment, and users, through monitoring, can determine if an equipment malfunction is due to a disturbance or problems within the equipment itself. Not surprisingly, standards have been introduced to cover this field. They define the types and sizes of disturbance, and the tolerance of various types of equipment to the possible

disturbances that may be encountered. The principal standards in this field are IEC 61000, EN 50160, and IEEE 1159. Standards are essential for manufacturers and users alike, to define what is reasonable in terms of disturbances that might occur and what equipment should withstand. Broad classifications of the disturbances that may occur on a power system are described as follows.

- Voltage Dips: The major cause of voltage dips on a system is local and remote faults, inductive loading, and switch on of large loads.
- Voltage surges: The major cause of Voltage surges on a system is Capacitor switching, Switch off of large loads and Phase faults.
- Overvoltage: The major cause of overvoltage on a system is Load switching, Capacitor switching, and System voltage regulation.
- Harmonics: The major cause of Harmonics on a system is Industrial furnaces Non-linear loads Transformers/generators, and Rectifier equipment.
- Power frequency variation: The major cause of Power frequency variation on a system is Loss of generation and Extreme loading conditions.
- Voltage fluctuation: The major cause of Voltage fluctuation on a system is AC motor drives, Inter-harmonic current components, and Welding and arc furnaces.
- Rapid voltage change: The major cause of Rapid voltage change on a system is Motor starting, Transformer tap changing.
- Voltage imbalance: The major cause of Voltage imbalance on a system is unbalanced loads, and Unbalanced impedances.
- Short and long voltage interruptions: The major cause of Short and long voltage interruptions on a system are Power system faults, Equipment failures, Control malfunctions, and CB tripping.
- Undervoltage: The major cause of Undervoltage on a system are Heavy network loading, Loss of generation, Poor power factor, and Lack of var support.
- Transients: The major cause of Transients on a system are Lightning, Capacitive switching, Non-linear switching loads, System voltage regulation.

## 2. Feature extraction

There are several reasons for monitoring power quality. The most important reason is the economic damage produced by electromagnetic phenomena in critical process loads. Effects on equipment and process operations can include malfunctions, damage, process disruption and other anomalies (Baggini 2010; IEEE 1995).

Troubleshooting these problems requires measuring and analyzing power quality and that leads us to the importance of monitoring instruments in order to localize the problems and find solutions. The critical stage for any pattern recognition problem is the extraction of discriminative features from the raw observation data. The combination of several scalar features forms the feature vector. For the classification of power quality events, the appropriate characteristics should be extracted. These characteristics should be chosen to have low computation time and can separate disturbances with high precision. Therefore, the similarities between these characteristics should be few and the number of them must be low (Moravej et al., 2010; Moravej et al., 2011a). Digital signal processing, or signal processing in short, concerns the extraction of features and information from measured digital signals. A wide variety of signal-processing methods have been developed through

the years both from the theoretical point of view and from the application point of view for a wide range of signals. Some prevalent used signal processing techniques are as follows (Moravej et al., 2010; Moravej et al., 2011a, 2001b; Heydt et al., 1999; Aiello 2005; Mishra et al., 2008):

- Fourier transform
- Short time Fourier transform
- Wavelet transform
- Hilbert transform
- Chirp Z transform
- S transform

More frequently, features extracted from the signals are used as the input of a classification system instead of the signal waveform itself, as this usually leads to a much smaller system input. Selecting a proper set of features is thus an important step toward successful classification. It is desirable that the selected set of features may characterize and distinguish different classes of power system disturbances. This can roughly be described as selecting features with a large interclass (or between-class) mean distance and a small intraclass (or within-class) distance. Furthermore, it is desirable that the selected features are uncorrelated and that the total number of features is small. Other issues that could be taken into account include mathematical definability, numerical stability, insensitivity to noise, invariability to affine transformations, and physical interpretability (Bollen & GU 2006). The signal decomposition and various parametric models of signals, where the extraction of signal characteristics (or features, attributes) becomes easier in some analysis domain as compared with directly using signal waveforms in the time domain.

### 3. Classification

These features can be used as the input of a classification system. For many real-world problems we may have very little knowledge about the characteristics of events and some incomplete knowledge of the systems. Hence, learning from data is often a practical way of analyzing the power system disturbances. Among numerous methodologies of machine learning, we shall concentrate on a few methods that are commonly used or are potentially very useful for power system disturbance analysis and diagnostics. These methods include:

(a) Learning machines using linear discriminates, (b) probability distribution-based Bayesian classifiers and Neyman-Pearson hypothesis tests, (c) multilayer neural networks, (d) statistical learning theory-based support vector machines, and (e) rule-based expert systems. A typical pattern recognition system consists of the following steps: feature extraction and optimization, selection of classifier topologies (or architecture), supervised/unsupervised learning, and finally the testing. It is worth noting that for each particular system some of these steps may be combined or replaced (Bollen & GU 2006).

Mathematically, designing a classifier is equivalent to finding a mapping function between the input space and the feature space followed by applying some decision rules that map the feature space onto the decision space. There are many different ways of categorizing the classifiers. A possible distinction is between statistical-based classifiers (e.g., Bayes, Neyman-Pearson, SVMs) and non-statistical-based classifiers (e.g., syntactic pattern recognition methods, expert systems). A classifier can be linear or nonlinear. The mapping

function can be parametric (e.g., kernel functions) or nonparametric (e.g., K nearest neighbors) as well. Some frequently used classifiers are listed below:

- Linear machines
- Artificial neural networks
- Bayesian classifiers
- Kernel machines
- Support vector machines
- Rule-based expert systems
- Fuzzy classification systems

## 4. Signal processing techniques

### 4.1 Short Time Fourier Transform

The purpose of the signal processing in various applications is to extract useful information through suitable technique. The non-stationary signals, in time-frequency analysis, need a tool that transform a signal into time frequency domain (Gu et al., 2000; Rioul & Vetterli 1991).

Fourier transform of  $x(t)w(t-\tau)$  is defined Short Time Fourier Transform (STFT). It relies on the assumption that the non-stationary signal  $x(t)$  can be segmented into sections confined by a window boundary  $w(t)$  within which it can be treated as the stationary one.

$$X_w(j\omega, \tau) = \int_{-\infty}^{+\infty} x(t)w(t-\tau)e^{-j\omega t} dt \quad (1)$$

$$\text{where: } w(t) = \begin{cases} 0 & t < 0, t > t_w \\ w(t) & 0 < t < t_w \end{cases}$$

The STFT depicts time evolution of the signal spectrum by shifted  $w(t)$  in time:  $w(t-\tau)$ . The discrete form of STFT can be expressed by following equation:

$$X_w(n, m) = \sum_{k=0}^{M-1} x(k)w(k-m)e^{-j\omega k} \quad (2)$$

where:

$\omega$  : an angle between sample,  $M$  : number of samples within the window.

The STFT as time-frequency analysis technique depends critically on the choice of the window. When a window has been selected for the STFT, the frequency resolution is unique at all frequency (Rioul & Vetterli 1991).

In (Gu et al., 2000) the spectral content as a function of time by using discrete STFT is obtained. Discrete STFT detects and analyses transients in the voltage disturbances by suitable selection of window size. Since the STFT has a fixed resolution at all frequency the interpretation of it terms of harmonics are easier. The band-pass filter outputs from discrete STFT are well associated with harmonics and are suitable for power system analysis. Also the STFT method is compared to wavelet in (Gu et al., 2000). The Authors of (Gu et al., 2000) believes that the choice of these methods depends heavily on particular applications. Overall it appears more favorable to use discrete STFT than dyadic wavelet and Binary-Tree Wavelet Filters (BT-WF) for voltage disturbance analysis.

## 4.2 Discrete Wavelet Transform (DWT)

Wavelet-based techniques are powerful mathematical tools for digital signal processing, and have become more and more popular since the 1980s. It finds applications in different areas of engineering due to its ability to analyze the local discontinuities of signals. The main advantages of wavelets is that they have a varying window size, being wide for slow frequencies and narrow for the fast ones, thus leading to an optimal time-frequency resolution in all the frequency ranges (Rioul & Vetterli 1991).

The DWT of a signal  $x$  is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response  $g$  resulting in a convolution of the two:

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n-k]. \quad (3)$$

The signal is also decomposed simultaneously using a high-pass filter  $h$ . The outputs giving the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass). It is important that the two filters are related to each other and they are known as a quadrature mirror filter. However, since half the frequencies of the signal have now been removed, half the samples can be discarded according to Nyquist's rule. The filter outputs are:

$$y_{\text{low}}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k] \quad (4)$$

$$y_{\text{high}}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n+1-k]. \quad (5)$$

This decomposition has halved the time resolution since only half of each filter output characterizes the signal. However, each output has half the frequency band of the input so the frequency resolution has been doubled. For multi level resolution the decomposition is repeated to further increase the frequency resolution and the approximation coefficients decomposed with high and low pass filters. This is represented as a binary tree with nodes representing a sub-space with different time-frequency localization. The tree is known as a filter bank (Moravej et al., 2010; Moravej et al., 2011a; Rioul & Vetterli 1991).

## 4.3 The Discrete S-Transform

The short term Fourier transforms (STFT) is commonly used in time-frequency signal processing (Stockwell 1991; Stockwell & Mansinha 1996). However, one of its drawbacks is the fixed width and height of the analyzing window. This causes misinterpretation of signal components with period longer than the window width; also the finite width limits time resolution of high-frequency signal components. One solution is to scale the dimensions of the analyzing window to accommodate a similar number of cycles for each spectral component, as in wavelets. This leads to the S-transform introduced by Stockwell, Mansinha and Lowe (Stockwell & Mansinha 1996). Like the STFT, it is a time-localized Fourier spectrum which maintains the absolute phase of each localized frequency component. Unlike the STFT, though, the S-transform has a window whose height and width frequency-varying (Stockwell 1991).



The S-transform was originally defined with a Gaussian window whose standard deviation is scaled to be equal to one wavelength of the complex Fourier spectrum. The original expression of S-transform as presented in (Stockwell 1991; Stockwell & Mansinha 1996) is

$$S(\tau, f) = \int_{-\infty}^{\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(\tau-t)^2 f^2}{2}} e^{-i2\pi f t} dt. \quad (6)$$

The normalizing factor of  $|f|/\sqrt{2\pi}$  in (1) ensures that, when integrated over all  $\tau$ ,  $S(\tau, f)$  converges to  $X(f)$ , the Fourier transform of  $x(t)$

$$\int_{-\infty}^{\infty} S(\tau, f) d\tau = \int_{-\infty}^{\infty} x(t) e^{-i2\pi f t} dt = X(f) \quad (7)$$

It is clear from (2) that  $x(t)$  can be obtained from  $S(\tau, f)$ . Therefore, S-transform is invertible. Let  $x[kT]$ ,  $k = 0, 1, \dots, N-1$  denote a discrete time series, corresponding to  $x(t)$ , with a time sampling interval of  $T$ . The discrete Fourier transform is given by

$$X\left[\frac{n}{NT}\right] = \frac{1}{N} \sum_{k=0}^{N-1} x[kT] e^{\frac{i2\pi n k}{N}} \quad (8)$$

where  $N = 0, 1, \dots, N-1$ . In the discrete case, the S-transform is the projection of the vector defined by the time series  $x[kT]$  onto a spanning set of vectors. The spanning vectors are not orthogonal, and the elements of the S-transform are not independent. Each basis vectors (of the Fourier transform) is divided into  $N$  localized vectors by an element-by-element product with  $N$  shifted Gaussians, such that the sum of these  $N$  localized vectors is the original basis vector.

Letting  $f \rightarrow n/NT$  and  $\tau \rightarrow jT$ , the discrete version of the S-transform is given in (Stockwell & Mansinha 1996) as follows:

$$S\left[jT, \frac{n}{NT}\right] = \sum_{m=0}^{N-1} X\left[\frac{m+n}{NT}\right] e^{\frac{2\pi^2 m^2}{n^2}} e^{\frac{i2\pi m j}{N}} \quad (9)$$

and for the  $n = 0$  voice is equal to the constant defined as

$$S[jT, 0] = \frac{1}{N} \sum_{m=0}^{N-1} x\left(\frac{m}{NT}\right) \quad (10)$$

where  $j, m$  and  $n = 0, 1, \dots, N-1$ . Equation (5) puts the constant average of the time series into the zero frequency voice, thus assuring the inverse is exact for the general time series.

Besides, standard ST suffers from poor energy concentration in the time-frequency domain. It gives degradation in time resolution at lower frequency and poor frequency resolution at higher frequency. In (Sahu et al., 2009) a modified Gaussian window has been proposed which scales with the frequency in an efficient manner to provide improved energy concentration of the S-transform. In this improved ST scheme the window function has been considered as the same Gaussian window but, an additional parameter  $\delta$  is introduced into the Gaussian window where its width varies with frequency as follows (Sahu et al., 2009):

$$\sigma(f) = \frac{\delta}{|f|}. \quad (11)$$

Hence the generalized ST becomes

$$S(\tau, f, \delta) = \int_{-\infty}^{+\infty} x(t) \frac{|f|}{\sqrt{2\pi\delta}} e^{-\frac{(\tau-t)^2 f^2}{2\delta^2}} e^{-j2\pi f t} dt \quad (12)$$

where the Gaussian window becomes

$$g(t, f, \delta) = \frac{|f|}{\sqrt{2\pi\delta}} e^{-\frac{t^2 f^2}{2\delta^2}} \quad (13)$$

and its frequency domain representation is

$$G(\alpha, f, \delta) = e^{-\frac{2\pi^2 \alpha^2 \delta^2}{f^2}} \quad (14)$$

The adjustable parameter  $\delta$  represents the number of periods of Fourier sinusoid that are contained within one standard deviation of the Gaussian window. If  $\delta$  is too small the Gaussian window retains very few cycles of the sinusoid. Hence the frequency resolution degrades at higher frequencies. If  $\delta$  is too high the window retains more sinusoids within it. As a result the time resolution degrades at lower frequencies. It indicates that the  $\delta$  value should be varied judiciously so that it would give better energy distribution in the time-frequency plane. The trade off between the time-frequency resolutions can be reduced by optimally varying the window width with the parameter  $\delta$ . At lower  $\delta$  value ( $\delta < 1$ ) the window widens more with less sinusoids within it, thereby it catches the low frequency components effectively. At higher  $\delta$  value ( $\delta > 1$ ) the window width decreases more with more sinusoids within it, thereby it resolves the high frequency components better. The parameter  $\delta$  is varied linearly with frequency within a certain range as given by (Stockwell 1991):

$$\delta(f) = k f \quad (15)$$

where  $k$  is the slope of the linear curve. The discrete version of (6) is used to compute the discrete IST by taking the advantage of the efficiency of the Fast Fourier Transform (FFT) and the convolution theorem.

#### 4.4 Hyperbolic S-transform

The S-transform has an advantage that it provides multi-resolution analysis while retaining the absolute phase of each frequency. But the Gaussian window has no parameter to allow its width to be adjusted in the time or frequency domain. Hence, the generalized ST which has a greater control over the window function has been introduced in (Pinnega & Mansinha 2003). Thus, at high frequencies, where the window is narrow and time resolution is good in any case, a more symmetrical window should be used. At low frequencies, where the window is wider and frequency resolution is less critical, a more asymmetrical window may be used to prevent the event from appearing too far ahead on the S-transform. This



concept led us to design the “hyperbolic” window  $w_{HY}$ . The hyperbolic window is a pseudo-Gaussian, obtained from the generalized window as follows (Pinnega & Mansinha 2003):

$$w_{HY} = \frac{2|f|}{\sqrt{2\pi}(\gamma_{HY}^F + \gamma_{HY}^B)} \times \exp \left\{ \frac{-f^2 \left[ X(\tau - t, \{\gamma_{HY}^F, \gamma_{HY}^B, \lambda_{HY}^2\}) \right]^2}{2} \right\} \quad (16)$$

where

$$X(\tau - t, \{\gamma_{HY}^F, \gamma_{HY}^B, \lambda_{HY}^2\}) = \frac{\gamma_{HY}^F + \gamma_{HY}^B}{2\gamma_{HY}^F \gamma_{HY}^B} (\tau - t - \zeta) + \frac{\gamma_{HY}^F - \gamma_{HY}^B}{2\gamma_{HY}^F \gamma_{HY}^B} \sqrt{(\tau - t - \zeta)^2 + \lambda_{HY}^2} \quad (17)$$

In equation (4),  $X$  is a hyperbola in  $(\tau - t)$  which depends upon a backward-taper parameter  $\gamma_{HY}^B$ , a forward-taper parameter  $\gamma_{HY}^F$  (we assume  $0 < \gamma_{HY}^F < \gamma_{HY}^B$ ), and a positive curvature parameter  $\lambda_{HY}$  which has units of time. The translation by  $\zeta$  ensures that the peak of  $w_{HY}$  occurs at  $(\tau - t) = 0$ ;  $\zeta$  is defined by

$$\zeta = \sqrt{\frac{(\gamma_{HY}^B - \gamma_{HY}^F)^2 \lambda_{HY}^2}{4\gamma_{HY}^B \gamma_{HY}^F}} \quad (18)$$

The output of HST is a  $N \times M$  matrix with complex values and is called the HS-Matrix whose rows pertain to frequency and whose columns pertain to time. Important information in terms of magnitude, phase and frequency can be extracted from the S-matrix.

Feature extraction is done by applying standard statistical techniques to the S-matrix. Many features such as amplitude, slope (or gradient) of amplitude, time of occurrence, mean, standard deviation and energy of the transformed signal are widely used for proper classification. Some extracted features based on S-transform are (Mishra et al., 2008; Gargoom et al., 2008).

- Standard deviation of magnitude contour.
- Energy of the magnitude contour.
- Standard deviation of the frequency contour.
- Energy of the of the frequency contour.
- Standard deviation of phase contour.

#### 4.5 Slantlet Transform

Selesnick in (Selesnick 1999) has proposed new transform method called Slantlet Transform (SLT) in signal processing application. The SLT primarily based on an improved model of Discrete Wavelet Transform (DWT) has evolved. The DWT utilize tree structure for implementation whereas the SLT filter-bank is implemented in type of a parallel structure with shorter support of component filters.

For better comparison between DWT and SLT, let two-scale orthogonal DWT namely  $D_2$  (which is proposed by Daubechies) and also the corresponding filter-bank actualized by SLT. It is potential to implement filters of shorter length whereas satisfying orthogonality and zero moment conditions (Selesnick 1999; Panda et al., 2002).

Recently, the SLT with improved time localization compared to DWT, is applicable in signal processing and data compression. In (Panda et al., 2002), the SLT used as a tool in making an effective method for compression of power quality disturbances. Data compression and reconstruction of impulse, sag, swell, harmonics, interruption, oscillatory transient and voltage flicker by using two-scale SLT was implemented.

Transforming of input signal by SLT, thresholding of transformed coefficients and reconstruction by inverse SLT are three main step of proposed method. Input data are generated using MATLAB code at sampling rate of 3 kHz. The obtained results, in (Panda et al., 2002), indicate SLT based compressing algorithm achieve better performance compared to Discrete Cosine Transform (DCT) and DWT.

Power quality events classification system contains SLT and fuzzy logic based classifier proposed in (Meher 2008). The decomposition level using SLT was selected as 3. five feature from SLT coefficient are selected as attributes to classification of disturbances. Six power quality events: sag, swell, interruption, transients, notch and spike are generated by MATLAB code at a sampling frequency of 5 kHz and up to 10 cycles. Under noisy conditions the proposed method is evaluated.

In (Hsieh et al., 2010), authors has implemented field programmable gate array (FPGA)-based hardware realization for identification of electrical power system disturbances. For each cycles of voltage interruption, sag, swell, harmonics and transient, 2950 points sampled and examined.

#### 4.6 Hilbert Transform

The Hilbert Transform (HT) is a signal processing method technique which is a linear operator in the mathematics. The HT of a signal  $x(t)$ :  $H[x(t)]$  is defined as (Kschischang 2006):

$$H[x(t)] = x(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(t - \tau)}{\tau} d\tau \quad (19)$$

where  $\tau$  is the shifting operator. The HT can be considered as the convolution of  $x(t)$  with the signal  $\frac{1}{\pi t}$ .

Clearly the HT of a signal  $x(t)$  in a time domain is another time domain signal  $H[x(t)]$ . The output of the HT is 90 degree phase shift of the original signal (Shukla et al., 2009).

A pattern recognition system has been proposed in (Jayasree et al., 2010) based on HT. The envelope of the power quality disturbances are calculated by using HT. The type of the power quality events is detected by the shape of the envelope. Some statistical information from the coefficients of HT is used. Means, standard deviation, peak value and energy of the HT coefficient are employed as input vector of the neural network classifier. Data generation is done by MATLAB/simulink. Sag, swell, transients, harmonics and voltage fluctuation along with normal signal were considered in (Jayasree et al., 2010). 500 samples from each class which were sample at 256 point/cycle with normal frequency 50 Hz. The method of (Jayasree et al., 2010) shows the HT features less sensitive to noise level.

The combination of EMD and HT are suggested in (Shukla et al., 2009). The HT is applied to first three IMF extracted from EMD to assess instantaneous amplitude and phase which are then employed for feature vector construction. The pattern recognition system used PNN classifier for identification the various power quality events.

Nine types of power quality disturbances are generated in MATLAB with a sampling frequency 3.2 kHz. The events are as follows: sag, swell, harmonics, flicker, notch, spike, transients, sag with harmonics, and swell with harmonics. Moreover, the efficiency of the method is compared to ST.

#### 4.7 Empirical Mode Decomposition

The Empirical Mode Decomposition (EMD) is a key part of Hilbert-Huang transform. The EMD is a technique for generating Intrinsic Mode Function (IMF) from non-linear and non-stationary signals.

An IMF which is belonging to a collection of IMF must be satisfied the following conditions (Huang et al., 1998; Rilling et al., 2003; Flandrin et al., 2004):

1. In the raw signal, the number of extrema and the number of zero-crossing must be equal (or extremely one number difference)
2. At any point, the mean value of the envelope specified by the local maxima and the envelope specified by the local minima is zero.

The decomposition procedure of EMD as called “sifting process”. The sifting process is used to extract IMF. The following steps to generate the IMF for signal  $X(t)$ , should be run:

- Determination of all extrema of raw signal  $X(t)$
- Using interpolation of local maxima and minima
- Calculation  $m_1$  as the mean of upper and lower envelope which are obtained from previous step
- Calculation of first component  $h_1$ :  $h_1 = X(t) - m_1$
- $h_1$  is treated as the raw data, then, by same process to obtain  $m_{11}$ ,  $m_{11}$  is also calculated and  $h_{11} = h_1 - m_{11}$
- The sifting process is repeated  $k$  times, until  $h_{1k}$  is an IMF, that is:  $h_{1(k-1)} - m_{1k} = h_{1k}$
- It is considered as the first IMF from the raw data  $c_1 = h_{1k}$
- The first extracted IMF component contains the shortest period variation in the data.

For the stopping of the sifting process, the residual,  $r$ , can be defined by  $X(t) - c_1 = r_1$ .

The residue,  $r_1$  treated as a new raw data and consequently same sifting procedure, as explained above, is applied. This algorithm can be repeated on  $r_j$ :  $r_{n-1} - c_n = r_n$  until all  $r_j$  are obtained. Last  $r_j$  can be calculated when the more IMF cannot be extracted in fact  $r_n$  becomes a monotonic function. Finally, From the above equations it is obtained:

$$X(t) = \sum_{i=1}^n c_i + r_i.$$

Therefore, a decomposition of the data into  $n$ -EMD modes is obtained (Huang et al., 1998).

In a few paper the EMD is applied for identification of different power quality events. One of the main reasons is the EMD method is relatively new. The high speed of the EMD is the most important advantage of the algorithm because all procedure is done in time domain only.

In (Lu et al., 2005) EMD proposed as a signal processing tool for power quality monitoring. The EMD has the ability to detect and localize transient features of power disturbances. In order to evaluate the proposed method, periodical notch, voltage dip, and oscillatory transient have been investigated. Three calculated IMF,  $c_1$ ,  $c_2$  and  $c_3$ , show the ability of the EMD to extract important features from several types of power quality disturbances. The

sampling frequency used in (Lu et al., 2005) is 5 kHz and the noisy condition, where signal to noise ratio value is about 26 dB, is considered.

## 5. Classification techniques

### 5.1 Feed Forward Neural Networks (FFNN)

Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the connections between elements largely determine the network function. Neural networks can be trained to perform a particular function by adjusting the values of the connections (weights and biases) between elements. Generally, neural networks are adjusted, or trained, so that a particular input leads desired target output. The network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Usually, many such input/target pairs are needed to train a network. Neural networks have been trained to perform complex functions in various fields, including pattern recognition, identification, classification, speech, vision, and control systems.

Neural networks can also be trained to solve problems that are difficult for conventional computers or human beings. Neural networks are usually applied for one of the three following goals:

- Training a neural network to fit a function
- Training a neural network to recognize patterns
- Training a neural network to cluster data

The training process requires a set of examples of proper network behavior i.e. network inputs and target outputs. During training the weights and biases of the network are iteratively adjusted to minimize the network performance function (Moravej et al., 2002).

### 5.2 Radial Basis Function Network (RBFN)

The RBFN model (Mao et al., 2000) consists of three layers: the inputs and hidden and output layers. The input space can either be normalized or an actual representation can be used. This is then fed to the associative cells of the hidden layer, which acts as a transfer function. The hidden layer consists of radial basis function like a sigmoidal function used in MLP network. The output layer is a linear layer. The RBF is similar to Gaussian density function, which is defined by the “center” position and “width” parameter. The RBF gives the maximum output when the input to the neuron is at the center and the output decreases away from the center. The width parameter determines the rate of decrease of the function as the input pattern distance increases from the center position. Each hidden neuron receives as net input the distance between its weight vector and the input vector. The output of the RBF layer is given as

$$O_k = \exp(-[X - C_k]^T [X - C_k] / 2\sigma_k^2) \quad (20)$$

where  $K = 1, 2, \dots, N$ , where  $N$  is the number of hidden nodes

$O_k$  = output of the  $K$  th node of the hidden layer

$X$  = input pattern vector

$C_k$  = center of the RBF of  $K$  th node of the hidden layer

$\sigma_k$  = spread of the  $K$  th RBF

Each neuron in the hidden layer outputs a value depending on its weight from the center of the RBF. The RBFN uses a Gaussian transfer function in the hidden layer and linear function in the output layer. The output of the  $j$  th node of the linear layer is given by

$$Y_j = W_j^T O_j \quad (21)$$

where  $j = 1, 2, \dots, M$ , where  $M$  is the number of output nodes

$Y_j$  = output of the  $j$  th node

$W_j$  = weight vector for node  $j$

$O_j$  = output vector from the  $j$  th hidden layer (can be augmented with bias vector)

Choosing the spread of the RBF depends on the pattern to be classified. The learning process undertaken by a RBF network may be visualized as follows. The linear weights associated with the output units of the network tend to evolve on a different “time scale” compared to the nonlinear activation functions of the hidden units. Thus, as the hidden layer’s activation functions evolve slowly in accordance with some nonlinear optimization strategy, the output layer’s weights adjust themselves rapidly through a linear optimization strategy. The important point to note is that the different layers of an RBF network perform different tasks, and so it is reasonable to separate the optimization of the hidden and output layers of the network by using different techniques, and perhaps operating on different time scales. There are different learning strategies that can be followed in the design of an RBF network, depending on how the centers of the radial basis functions of the network are specified. Essentially following three approaches are in use:

- Fixed centers selected at random
- Self-organized selection of centers
- Supervised selection of centers

### 5.3 Probabilistic neural network (PNN)

The PNN was first proposed in (Spetch 1990; Mao et al., 2000). The development of PNN relies on the Parzen window concept of multivariate probability estimates. The PNN combines the Baye’s strategy for decision-making with a non-parametric estimator for obtaining the Probability Density Function (PDF) (Spetch 1990; Mao et al., 2000). The PNN architecture includes four layers; input, pattern, summation, and output layers. The input nodes are the set of measurements. The second layer consists of the Gaussian functions formed using the given set of data points as centers. The third layer performs an average operation of the outputs from the second layer for each class. The fourth layer performs a vote, selecting the largest value. The associated class label is then determined (Spetch 1990). The input layer unit does not perform any computation and simply distributes the input to the neurons. The most important advantages of PNN classifier are as below:

- Training process is very fast
- An inherent parallel structure
- It converges to an optimal classifier as the size of the representative training set increases
- There are not local minima issues
- Training patterns can be added or removed without extensive retraining



#### 5.4 Support vector machines (SVMs)

The SVM finds an optimal separating hyperplane by maximizing the margin between the separating hyperplane and the data (Cortes et al., 1995; Vapnik 1998; Steinwart 2008, Moravej et al., 2009). Suppose a set of data  $T = \{x_i, y_i\}_{i=1}^m$  where  $x_i \in \mathbb{R}^n$  denotes the feature vectors,  $y_i \in \{+1, -1\}$  stands for two classes, and  $m$  is the sample number, if two classes are linearly separable, the hyperplane  $f(x) = 0$  can be determined such that separates the given feature vectors.

$$f(x) = w \cdot x + b = \sum_{k=1}^m w_k \cdot x_k + b = 0 \quad (22)$$

where  $w$  and  $b$  denote the weight vector and the bias term, respectively. The position of the separating hyperplane is defined by setting these parameters. Thus the separating hyperplane satisfy the following constraints:

$$y_i f(x_i) = y_i (w \cdot x_i + b) \geq 1, \quad i = 1, 2, \dots, m \quad (23)$$

$\xi_i$  are positive slack variables that measure the distance between the margin and the vectors  $x_i$  that lie on the incorrect side of the margin. Then, in order to obtain the optimal hyperplane, the following optimization problem must be solved:

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i, \quad i = 1, 2, \dots, m \\ &\text{subject to} \quad \begin{cases} y_i (w \cdot x_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \end{aligned} \quad (24)$$

where  $C$  is the error penalty.

By introducing the Lagrangian multipliers  $\alpha_i$ , the above-mentioned optimization problem is transformed into the dual quadratic optimization problem, as follows:

$$\text{Maximize} \quad L(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (25)$$

$$\text{Subject to} \quad \sum_{i=1}^m \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, 2, \dots, m \quad (26)$$

Thus, the linear decision function is created by solving the dual optimization problem, which is defined as:

$$f(x) = \text{sign} \left( \sum_{i,j=1}^m \alpha_i y_i (x_i \cdot x_j) + b \right) \quad (27)$$

If the linear classification is not possible, the nonlinear mapping function  $\phi$  can be used to map the original data  $x$  into a high dimensional feature space that the linear classification is possible. Then, the nonlinear decision function is:

$$f(x) = \text{sign} \left( \sum_{i,j=1}^m \alpha_i y_i K(x_i, x_j) + b \right) \quad (28)$$



where  $K(x_i, x_j)$  is called the kernel function,  $K(x_i, x_j) = \phi(x_i)\phi(x_j)$ . Linear, polynomial, Gaussian radial basis function and sigmoid are the most commonly used kernel functions (Cortes et al., 1995; Vapnik 1998; Steinwart 2008). To classify more than two classes, two simple approaches could be applied. The first approach uses a class by class comparison technique with several machines and combines the outputs using some decision rule. The second approach for solving the multiple class detection problem using SVMs is to compare one class against all others, therefore, the number of machines is the same number of classes. These two methods have been described in details in (Steinwart 2008).

### 5.5 Relevance Vector Machines

Michael E. Tipping proposed Relevance Vector Machine (RVM) in 2001 (Tipping 2000). It assumes knowledge of probability in the areas of Bayes' theorem and Gaussian distributions including marginal and conditional Gaussian distributions (Fletcher 2010). RVMs are established upon a Bayesian formulation of a linear model with an appropriate prior that cause a sparse representation. Consequently, they can generalize well and provide inferences at low computational cost (Tzikas 2006; Tipping 2000). The main formulation of RVMs is presented in (Tipping 2000).

New combination of WT and RVMs are suggested in (Moravej et al., 2011a) for automatic classification of power quality events. The Authors in (Moravej et al., 2011a) employed the WT techniques to extract the feature from details and approximation waves. The constructed feature vectors as input of RVM classifier are applied for training the machines to monitoring the power quality events. The feature extracted from various power quality signals are as follow:

1. Standard deviation of level 2 of detail.
2. Minimum value of absolute of level  $n$  of approximation. ( $n$  is desirable decomposition levels)
3. Mean of average of absolute of all level of details.
4. Mean of disturbances energy.
5. Energy of level 3 of detail.
6. RMS value of main signal.

Sag, swell, interruption, harmonics, swell with harmonics, sag with harmonics, transient, and flicker, was studied. Data is generated by parametric equation an MATLAB environment. The procedure of classification is tested in noisy conditions and the results show the efficiency of the method. The CVC method for classification of nine power quality events is proposed. First time that RVM based classifier for recognition of power quality events is applied in (Moravej et al., 2011a).

### 5.6 Logistic Model Tree and Decision Tree

Logistic Model Tree (LMT) is a machine for supervised learning issues. The LMT combines linear logistic regression and tree induction. The LogitBoost algorithm for building the structure of logistic regression functions at the nodes of a tree is used. Also, the renowned Classification and Regression Tree (ACRT) algorithm for pruning are employed. The LogitBoost is employed to pick the foremost relevant attributes in the data when performing logistic regression by performing a simple regression in each iteration and stopping before convergence to the maximum likelihood solution. The LMT does not require any tuning of parameters by the user (Landwehr 2005; Moravej et al., 2011b).

A LMT includes standard Decision Tree (DT) (Kohavi & Quinlan 1999) structure with logistic regression functions at the leaves. Compared to ordinary DTs, a test on one of the attributes is related to every inner node.

The new combination as pattern recognition system has been proposed in (Moravej et al., 2011b). The Authors used LMT based classifier for identification of nine power quality disturbances. Sag, swell, interruption, harmonics, transient, and flicker, was studied. Simultaneously disturbances consisting of sag and harmonics, as well as swell and harmonics, are also considered. Data is generated by parametric equation in MATLAB environment. The sampling frequency is 3.2 kHz. The feature vector composed of four features extracted by ST method. In (Moravej et al., 2011b), the features are based on the Standard Deviation (SD) and energy of the transformed signal and are extracted as follows:

Feature 1: SD of the dataset comprising the elements corresponding to the maximum magnitude of each column of the S-matrix.

Feature 2: Energy of the dataset comprising of the elements corresponding to the maximum magnitude of each column of the S-matrix.

Feature 3: SD of the dataset values corresponding to the maximum value of each row of the S-matrix.

Feature 4: Energy of the dataset values corresponding to the maximum value of each row of the S-matrix.

For classification of power quality disturbances, 100 cases of each class are generated for the training phase, and another 100 cases are generated for the testing phase (Moravej et al., 2011b). The Sensitivity of the algorithm, in (Moravej et al., 2011b), is also investigated under noisy condition.

## 6. Pattern recognition techniques

The functionality of an automated pattern recognition system can be divided into two basic tasks, as shown in Fig. 1: the description task generates attributes of PQ disturbances using feature extraction techniques, and the classification task assigns a group label to the PQ disturbance based on those attributes with a classifier. The description and classification tasks work together to determine the most accurate label for each unlabeled object analyzed by the pattern recognition system (Moravej et al., 2010; Moravej et al., 2011a).

Feature extraction is a critical stage because it reduces the dimension of input data to be handled by the classifier. The features which truly discriminate among groups will assist in identification, while the lack of such features can prevent the classification task from arriving at an accurate identification. The final result of the description task is a set of features, commonly called a feature vector, which constitutes a representation of the data. The classification task uses a classifier to map a feature vector to a group. Such a mapping can be specified by hand or, more commonly, a training phase is used to induce the mapping from a collection of feature vectors known to be the representative of the various groups among which discrimination is being performed (i.e., the training set). Once formulated, the mapping can be used to assign identification to each unlabeled feature vector subsequently presented to the classifier. So, it is obvious that a good feature extraction technique should be able to derive significant feature vectors in an automated way along with determining less number of relevant features to characterize the complete systems. Thus, the subsequent computational burden of the classifier can be reduced.

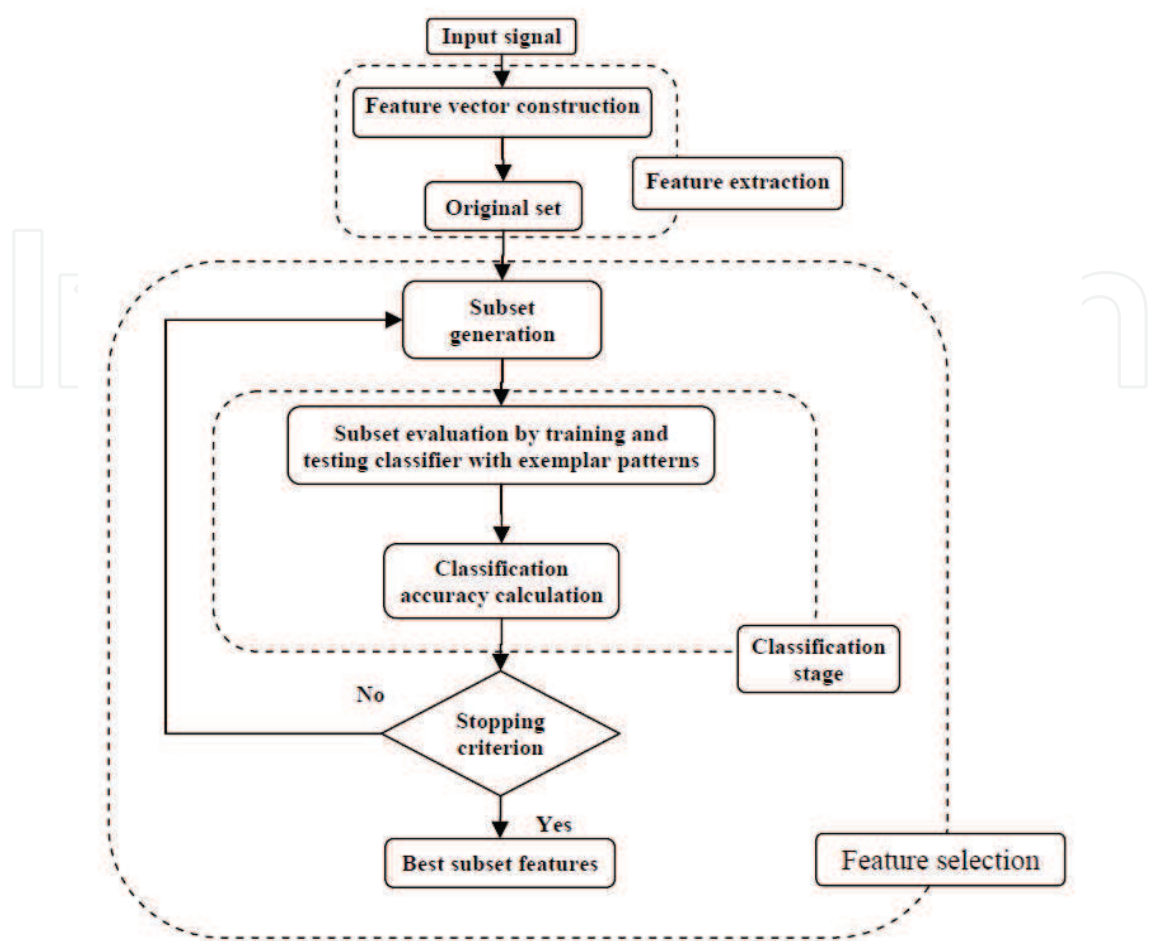


Fig. 1. General pattern recognition algorithm for PQ events classification

7. Feature selection

By removing the most irrelevant and redundant features from the data, feature selection helps to improve the performance of learning models by alleviating the effect of the curse of dimensionality, enhancing generalization capability, speeding up learning process and improving model interpretability. If the size of initial feature set is large, exhaustive search may not be feasible due to processing time considerations. In that case, a suboptimal selection algorithm is preferred. However, none of these algorithms guarantee that the best feature set is obtained. The selection methods provide useful information about superiority of selected features, superiority of feature selection strategy and the relation between the useful features and the desired feature size (Gunal et al., 2009). Generally the feature selection methods give answer to some question arises from PQ classification problem as follows.

7.1 Filter

Filter type methods are based on data processing or data filtering methods. Features are selected based on the intrinsic characteristics, which determine their relevance or discriminate powers with regard to the targeted classes. Some of these methods are described as follows (Proceedings of the Workshop on Feature Selection for Data Mining).

### 7.1.1 Correlation

A correlation function is the correlation between random variables at two different points in space or time, usually as a function of the spatial or temporal distance between the points. If one considers the correlation function between random variables representing the same quantity measured at two different points then this is often referred to as an autocorrelation function being made up of autocorrelations. Correlation functions of different random variables are sometimes called cross correlation functions to emphasize that different variables are being considered and because they are made up of cross correlations. Correlation functions are a useful indicator of dependencies as a function of distance in time or space, and they can be used to assess the distance required between sample points for the values to be effectively uncorrelated. In addition, they can form the basis of rules for interpolating values at points for which there are observations.

The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient, or "Pearson's correlation." It is obtained by dividing the covariance of the two variables by the product of their standard deviations. The population correlation coefficient  $\rho_{X,Y}$  between two random variables  $X$  and  $Y$  with expected values  $\mu_X$  and  $\mu_Y$  and standard deviations  $\sigma_X$  and  $\sigma_Y$  is defined as (Rodgers & Nicewander 1988; Dowdy & Wearden 1983):

$$\rho_{X,Y} = \text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (29)$$

where  $E$  is the expected value operator,  $\text{cov}$  means covariance, and,  $\text{corr}$  a widely used alternative notation for Pearson's correlation.

The Pearson correlation is +1 in the case of a perfect positive (increasing) linear relationship (correlation), -1 in the case of a perfect decreasing (negative) linear relationship (anticorrelation), and some value between -1 and 1 in all other cases, indicating the degree of linear dependence between the variables. As it approaches zero there is less of a relationship (closer to uncorrelated). The closer the coefficient is to either -1 or 1, the stronger the correlation between the variables. Some feature can be selected from feature space based on the obtained correlation coefficient of potential features.

### 7.1.2 Product-Moment Correlation Coefficient (PMCC)

For each signal, a set of features may be extracted that characterize the signal. The purpose of the feature selection is to reduce the dimension of feature vector while maintaining admissible classification accuracy. In order to, select the most meaningful features Product-Moment Correlation Coefficient (PMCC or Pearson correlation) method has been applied to feature vector obtained in feature extraction step.

The Pearson correlation between two variables  $X$  and  $Y$ , giving a value between +1 and -1. A correlation of +1 means that there is a perfect positive linear relationship between variables. A correlation of -1 means that there is a perfect negative linear relationship between variables. A correlation of 0 means there is no linear relationship between the two variables.

Pearson's correlation coefficient between two variables is computed as (Son & Baek 2008):

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_X S_Y} \quad (30)$$

where

$r$  : correlation coefficient

$\bar{X}, \bar{Y}$  : the means of  $X$  and  $Y$  respectively

$S_x, S_y$  : the standard deviation of  $X$  and  $Y$  respectively.

The correlation coefficient  $r$  is selected as 0.95, 0.9, and 0.85 respectively. The extracted features, those have correlation more than  $r$  will be removed automatically. The dimension reduction of the feature vector has several advantages including: low computational burden for training and testing phases of machine learning, high speed of training phase, and minimization of misclassifications.

Afterwards, feature normalization is applied to ensure that each feature in a feature vector is properly scaled. Therefore, the different features are equally weighted as an input of classifiers.

### 7.1.3 Minimum Redundancy Maximum Relevance (MRMR)

The MRMR method that considers the linear independency of the feature vectors as well as their relevance to the output variable so it can remove redundant information and collinear candidate inputs in addition to the irrelevant candidates. This technique is done in two steps. At first if the mutual information between a candidate variable and output feature is greater than a pre specified value then it is kept for further processing else it is discarded. This is the first step of the feature selection technique ("Maximum Relevance" part of the MRMR principle). In the next step, the cross-correlation is performed on the retained features obtained from the first step. If the correlation coefficient between two selected features is smaller than a pre specified value both features are retained; else, only the features with largest mutual information are retained. The cross-correlation process is the second step of the feature selection technique ("Minimum Redundancy" part of the MRMR principle) (Peng et al., 2005). So, the proposed feature selection technique is composed of a mutual information based filter to remove irrelevant candidate inputs and a correlation based filter to remove collinear and redundant candidates. Two thresholds values must be determined for two applied filters in the first and second steps. Retained variables after the two steps of the feature selection are selected as the input of the forecast engine. In order to obtain an efficient classification scheme, threshold values (adjustable parameters) must be fine tuned.

## 7.2 Wrapper

Wrapper based methods use a search algorithm to seek through the space of possible features and evaluate each subset by running a model on the selected subset. Wrappers usually need huge computational process and have a risk of over fitting to the model.

### 7.2.1 Sequential forward selection

Sequential forward selection was first proposed in (Whitney 1971). It operates in the bottom-to-top manner. The selection procedure begins with a null subset initially. Then, at each step, the feature that maximizes the criterion function is added to the current subset. This procedure continues until the requested number of features is selected. The nesting effect is present such that a feature added into the set in a step can not be removed in the subsequent steps (Gunal et al., 2009). As a consequence, SFS method can yield only the suboptimal result.



### 7.2.2 Sequential backward selection

Sequential backward selection method proposed in works in a top-to-bottom manner (Marill & Green 1963). It is the reverse case of SFS method. Initially, complete feature set is considered. At each step, a single feature is removed from the current set so that the criterion function is maximized for the remaining features within the set. The removal operation continues until the desired number of features is obtained. Once a feature is eliminated from the set, it can not enter into the set in the subsequent steps (Gunal et al., 2009).

### 7.2.3 Genetic algorithm

Genetic algorithms belong to the larger class of Evolutionary Algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as mutation, selection, and crossover. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population based on their fitness, and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population (Yang & Honavar 1998). The chromosomes are encoded with {0, 1} binary alphabet. In a chromosome, "1" indicates the selected features while "0" indicates the unselected ones. For example, a chromosome defined as:

$$\{1 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1\} \quad (31)$$

specifies that the features with index 1, 3, 5, 6, and 10 are selected while the others are unselected. The fitness value corresponding to a chromosome is usually defined as the classification accuracy obtained with the selected features.

### 7.2.4 Generalized sequential forward selection (GSFS)

In generalized version of SFS, instead of single feature,  $n$  features are added to the current feature set at each step (Kittler 1978). The nesting effect is still present.

### 7.2.5 Generalized sequential backward selection (GSBS)

In generalized form of SBS (GSBS), instead of single feature,  $n$  features are removed from the current feature set at each step. The nesting effect is present here, too (Kittler 1978).

### 7.2.6 Plus-1 takeaway-r (PTA)

The nesting effect present in SFS and SBS can be partly avoided by moving in the reverse direction of selection for certain number of steps. With this purpose, at each step,  $l$  features are selected using SFS and then  $r$  features are removed with SBS. This method is called as PTA (Stearn 1976). Although the nesting effect is reduced with respect to SFS and SBS, PTA still provides suboptimal results.

### 7.2.7 Sequential forward floating selection (SFFS)

Dynamic version of PTA leads to SFFS method. Unlike the PTA method that parameters  $l$  and  $r$  are fixed, they are float in each step (Pudil et al., 1994). Thus, sub-selection searching



process, different number of features can be added to or removed from the set until a better criterion value is attained. The flexible structure of SFSS causes the feature dimension to be different in each step.

8. Review of proposed pattern recognition algorithms and conclusions

In the Table 1 some references are mentioned which use the pattern recognition schemes for detection of power quality events. These detection algorithms are composed of combination a feature extraction and a classification method.

Reference	Method	Reference	Method
(Moravej et al., 2010; Eristi & Demir 2010)	WT+SVM	(Behera et al., 2010)	ST+Fuzzy
(Moravej et al., 2011a)	WT+RVM	(Huang et al., 2010)	HST+PNN
(Moravej et al., 2011b)	ST+LMT	(Meher 2008)	SLT+Fuzzy
(RathinaPrabha 2009; Hooshmand & Enshaee 2010)	DFT+Fuzzy	(Jayasree et al., 2010)	HT+RBFN
(Mehera et al., 2004; Kaewarsa et al., 2008)	WT+ANN	(Mishra et al., 2008)	ST+PNN
(Liao & Lee 2004; Hooshmand & Enshaee 2010)	WT+Fuzzy	(Zhang et al., 2011)	DFT+DT
(Uyar et al., 2009; Hooshmand & Enshaee 2010)	ST+ANN		

Table 1. Review of proposed pattern recognition algorithms

Power Quality is a term used to broadly encompass the entire scope of interaction among electrical suppliers, the environment, the systems and products energized, and the users of those systems and products. It is more than the delivery of "clean" electric power that complies with industry standards. It involves the maintainability of that power, the design, selection, and the installation of every piece of hardware and software in the electrical energy system.

Many algorithms have been proposed for detection and classification of power quality events. Pattern recognitions schemes are very popular solution for detection of power quality events. The combinations of signal processing and classification tools have been widely applied in detection methods. The most useful features are extracted by analysis of signals and then they are discriminated by using a classifier or by definition of a proper index.

9. References

Aiello M.; Cataliotti A.; Nuccio S (2005). A Chirp-Z transform-based synchronizer for power system measurements, *IEEE Transaction on Instrument Measurement*, Vol. 54, No. 3, (2005), pp. 1025-1032.

ATPDraw for Windows 3.1x/95/NT version 1.0 User’s Manual, Trondheim, Norway 15th October 1998.

- Baggini A. (2008), *Handbook of Power Quality*, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2008.
- Behera, H.S. Dash, P.K.; Biswal, B. Power quality time series data mining using S-transform and fuzzy expert system, (2010) *Applied Soft Computing*, Vol. 10, (2010), pp. 945–955.
- Bollen, M.H.J.; GU, Y.H. (2006) Signal Processing of Power Quality Disturbances, *Institute of Electrical and Electronics Engineers*, Published by John Wiley & Sons, Inc.
- Cortes, C, Vapnik, V. Support vector networks. *Machine Learning*, Vol. 20, pp. 273-297, 1995.
- Dowdy, S.; Wearden, S.; Statistics for Research, Wiley, 1983. ISBN 0471086029, pp. 230.
- Eristi, H.; Y. Demir, A new algorithm for automatic classification of power quality events based on wavelet transform and SVM, *Expert Systems with Applications*, Vol. 37, (2010), pp. 4094–4102.
- Flandrin, P.; Rilling, G.; Gonçalves P. (2004). Empirical mode decomposition as a filter bank, *IEEE Signal Processing Letters*, Vol.11, No.2, (February 2004), pp. 112-114.
- Fletcher, T. Relevance Vector Machines Explained, 2010, [www.cs.ucl.ac.uk/sta/T.Fletcher/](http://www.cs.ucl.ac.uk/sta/T.Fletcher/).
- Gaig Z.L (2004). Wavelet-based neural network for power disturbance recognition and classification. *IEEE Transaction on Power Delivery*, Vol.19, No.4, (2004), pp. 1560–1568.
- Gargoom, M.; Wen, N.E.; Soong, L. (2008). Automatic Classification and characterization of power quality events. *IEEE Transaction on Power Delivery*, Vol. 23, No. 4, (2008), pp. .
- Gu, Y.H.; M.H.J., Bollen (2000). Time-frequency and time-scale domain analysis of voltage disturbances, *IEEE Transaction on Power Delivery*, Vol. 15, No. 4, (October 2000), .
- Gunal, S.; Gerek, O.N.; Ece, D.G.; Edizkan, R (2009). The search for optimal feature set in power quality event classification. *Expert Systems with Application*, Vol. 36, (2009), pp. 10266–10273.
- Heydt, G.T.; Fjeld P.S.; Liu, C.C.; Pierce, D.; Tu, L.; Hensley, G. (1999) Applications of the windowed FFT to electric power quality assessment. *IEEE Transaction Power Delivery*, Vol. 14, No. 4, (1999), pp. 1411–1416.
- Hooshmand, R.; Enshaee, A.; (2010) Detection and classification of single and combined power quality disturbances using fuzzy systems oriented by particle swarm optimization algorithm, *Electric Power Systems Research*, Vol. 80, (2010), pp. 1552–1561.
- Hsieh, C.T.; Lin, J.M.; Huang, S.J. (2010). Slant transform applied to electric power quality detection with field programmable gate array design enhanced, *Electrical Power and Energy Systems*, Vol. 32, (2010) pp. 428–432.
- Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Yen, Q.Z.N. ; Tung, C.C.; Liu, H.H.; The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. Lond. A*, Printed in Great Britain (1998), 454, pp. 903-995.
- Huang, N.; Xu, D.; Liu, X. (2010), Power Quality Disturbances Recognition Based on HS-transform, *First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA)*, Issue Date: 17-19 Sept. 2010.
- IEEE 1159: 1995, Recommended practice for monitoring electric power quality, 1995.
- Jayasree, T.; Devaraj, D.; Sukanesh, R (2010). Power quality disturbance classification using Hilbert transform and RBF networks, *Neurocomputing*, Vol. 73, (2010), pp. 1451–1456.

- Kaewarsa, S.; Attakitmongkol, K.; Kulworawanichpong, T. (2008), Recognition of power quality events by using multi wavelet-based neural networks. *International journal of Electric Power Energy and System*, Vol. 30, (2008), pp. 254–260.
- Kittler, J. Feature set search algorithms. In C. H. Chert (Ed.), *Pattern recognition and signal processing*, 1978, pp. 41–60. Mphen aan den Rijn, Netherlands: Sijthoff and Noordhoff.
- Kohavi, R.; Quinlan, R.; *Decision Tree Discovery*, Data Mining, University of New South Wales, Sydney 2052 Australia, 1999.
- Kschischang, F.R.; *The Hilbert Transform*, Department of Electrical and Computer Engineering University of Toronto, October 22, 2006.
- Landwehr, N., Hall, M., and Frank, E., *Logistic Model Tree*, *Machine Learning*, Springer Science, Vol. 59, (2005) pp.161-205.
- Liao, Y.; Lee, J.B.; A fuzzy-expert system for classifying power quality disturbances, *Electrical Power and Energy Systems*, (2004), Vol. 26, pp. 199–205.
- Lu, Z.; Smith, J.S.; Wu, Q.H.; Fitch, J. (2005) Empirical mode decomposition for power quality monitoring, *2005 IEEE/PES Transmission and Distribution Conference & Exhibition: Asia and Pacific Dalian, China*.
- Mao, K.Z.; Tan, K.C.; Ser, W. Probabilistic neural-network structure determination for pattern classification, *IEEE Transaction on Neural Networks*, Vol. 11, (2000), pp. 1009–1016.
- Marill, T.; Green, D.M.; (1963), On the effectiveness of receptors in recognition systems. *IEEE Transaction on Information Theory*, Vol. 9, (1963), pp. 11–17.
- MATLAB 7.4 Version Wavelet Toolbox, Math Works Company, Natick, MA. (MATLAB)
- Meher, S.K. (2008), A Novel Power Quality Event Classification Using Slantlet Transform and Fuzzy Logic, (2008) *International Conference on Power System Technology India*, 2008.
- Mehera, S.K.; Pradhan, A.K.; Panda, G. (2004). An integrated data compression scheme for power quality events using spline wavelet and neural network. *Electric Power Systems Research*, Vol. 69, (2004), pp. 213–220.
- Mishra, S.; Bhende, C.N.; Panigrahi, B.K. (2008) Detection and classification of power quality disturbances using S-transform and probabilistic neural network, *IEEE Transaction on Power Delivery*, Vol. 23, No. 1, (January 2008), pp. 280–287.
- Moravej, Z.; Vishvakarma, D.N.; Singh, S.P.; (2002) Protection and condition monitoring of power transformer using ANN. *Electric Power Components and systems*, Vol.30, No.3, (March 2002), pp. 217–231.
- Moravej, Z.; Vishwakarma, D.N.; Singh, S.P. (2003) Application of radial basis function neural network for differential relaying of a power transformer, *Computers & Electrical Engineering*, Vol.29, No.3, (May 2003) pp. 421–434.
- Moravej, Z. ; Banihashemi, S.A.; Velayati, M.H. (2009), Power quality events classification and recognition using a novel support vector algorithm. *Energy Conversion and Management*, Vol. 50, (2009), pp. 3071–3077.
- Moravej Z.; Abdoos A.A.; Pazoki M. (2010). Detection and classification of power quality disturbances using wavelet transform and support vector machines. *Electric Power Components and Systems*, Vol.38, (2010), pp. 182–196.

- Moravej Z. Pazoki M.; Abdoos A.A (2011). Wavelet transform and multi-class relevance vector machines based recognition and classification of power quality disturbances, *European Transaction on Electrical Power*, Vol.21, No.1, (January 2011), pp. .
- Moravej, Z.; Abdoos, A.A.; Pazoki, M.; (2011), New Combined S-transform and Logistic Model Tree Technique for Recognition and Classification of Power Quality Disturbances, *Electric Power Components and Systems*, Vol.39, No.1, (2011), pp. 80-98.
- Oleskovicz, M.; Coury, D.V.; Felho, O.D.; Usida, W.F.; Carneiro, A.F.M.; Pires, R.S.; Power quality analysis applying a hybrid methodology with wavelet transforms and neural networks. *Electrical Power and Energy Systems*, Vol. 31, (2009), pp. 206-212.
- Panda, G.; Dash, P.K.; Pradhan, A.K.; & Meher S.K. (2002). Data Compression of Power Quality Events Using the Slantlet Transform, *IEEE Transaction on Power Delivery*, Vol.17, No.2, (April 2002), pp. 662-662.
- Peng, H.; Long, F.; Ding C.; (2005). Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.27, No.8, (August 2005), pp. 1226-1238.
- Pinnega, C.R.; Mansinha L.; The S -transform with windows of arbitrary and varying shape. *Geophysics*, Vol. 68, (2003), pp. 381-385.
- Proceedings of the Workshop on Feature Selection for Data Mining: Interfacing Machine Learning and Statistics in conjunction with the 2005 SIAM International Conference on Data Mining April 23, 2005 Newport Beach, CA.
- PSCAD/EMTDC Power systems simulation Manual, Winnipeg, MB, Canada, 1997.
- Pudil, P.; Novovicova, J.; Kittler, J. (1994), Floating search methods in feature selection. *Pattern Recognition Letters*, Vol. 15, (1994), pp. 1119-1125.
- Rathina Prabha, N.; Marimuthu, N.S.; Babulal, C.K. (2009). Adaptive neuro-fuzzy inference system based total demand distortion factor for power quality evaluation, *Neurocomputing*, Vol. 73, (2009), pp. 315-323.
- Rilling, G.; Flandrin, P.; Gonçalves, P. (2003).ON Empirical mode decomposition and its algorithm. (2003) *IEEE-EURA SIP Workshop on Nonlinear Signal and Image Processing NSIP-03*, Grado (I), 2003.
- Rioul, O.; Vetterli, M. (1991). Wavelet and Signal Processing. *IEEE Signal Processing Magazine*, Vol. 8, No. 4, (October 1991), pp. 14 - 38.
- Rodgers, J.L.; Nicewander, W.A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, Vol. 42, (1988), pp. 59-66.
- Sahu, S.S.; Panda G.; George, N.V. (2009). An improved S-transform for time-frequency analysis. *IEEE International Advance Computing Conference (IACC 2009)* Patiala, India, 6-7 March 2009, pp. 315-319.
- Selesnick, I.W. (1999), The Slantlet Transform, *IEEE Transaction On Signal Processing*, Vol.47, No.5, (May 1999), pp. 1304-1313.
- Shukla, S.; Mishra, S.; Singh, B. (2009). Empirical-mode decomposition with Hilbert transform for power-quality assessment, *IEEE Transaction on Power Delivery*, Vol. 24, No. 4, (October 2009), pp. 2159-2165.
- Son, Y.S.; Baek, J.; A modified correlation coefficient based similarity measure for clustering time-course gene expression data, *Pattern Recognition Letters*, Vol. 29, (2008) pp. 232-242.
- Specht, D.F.; Probabilistic neural network. *Neural networks*, Vol. 1, pp.109-118, 1990.



- Stearn, S.D.; On selecting features for pattern classifiers. In *Proceedings of the third international conference on pattern recognition*, 1976, pp. 71–75, Coronado, CA.
- Steinwart, A. Christmann. *Support Vector Machines*; New York: Springer, 2008.
- Stockwell, R.G (1991). Why use the S-transform?. Northwest Research Associates, Colorado Research Associates Division, 3380 Mitchell Lane, Boulder Colorado USA 80301. American Mathematical Society 1991.
- Stockwell, R.G.; Mansinha, L.; Lowe, R.P. (1996). Localization of the complex spectrum: the S-transform. *IEEE Transactions on Signal Processing*, (1996), Vol. 4, pp. 998–1001.
- Tipping, M.E. (2000), the relevance vector machine. Solla SA, Leen TK, Muller KR. *Advances in Neural Information Processing Systems 2000*; 12: MIT Press, Cambridge, MA.
- Tipping ME. The Relevance Vector Machine. Microsoft Research, U.K, 2000.
- Tzikas, G.D.; Wei, L.; Likas, A.; Yang, Y.; Galatsanos, N.P. (2006). A tutorial on relevance vector machines for regression and classification with applications, University of Ioannina, Ioanni, GREECE, Illinois Institute of Technology, Chicago, USA, 2006.
- Uyar, M.; Yildirim, S.; Gencoglu, M.T.; (2009). An expert system based on S-transform and neural network for automatic classification of power quality disturbances, *Expert Systems with Applications*, Vol. 36, (2009), pp. 5962–5975.
- Uyara, M.; Yildirima, S. & Gencoglu MT (2008). An effective wavelet-based feature extraction method for classification of power quality disturbance signals. *Electric Power System Research*, Vol. 78, (2008), pp. 1747–1755.
- Vapnik, V.N. *Statistical Learning Theory*; Wiley: New York, 1998.
- Whitney, A.W. (1971) A direct method of nonparametric measurement selection. *IEEE Transaction on Computers*, Vol.20, (1971), pp. 1100–1103.
- Yang, J.; Honavar, V. (1998), Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems and their Applications*, Vol. 2, (1998), pp. 44–49.
- Zhang, M.; Li, K.; Hu, Y.; (1998) A real-time classification method of power quality disturbances, *Electric Power Systems Research*, Vol. 81, (1998), pp. 660–666, 2011.

IntechOpen



## **Power Quality – Monitoring, Analysis and Enhancement**

Edited by Dr. Ahmed Zobaa

ISBN 978-953-307-330-9

Hard cover, 364 pages

**Publisher** InTech

**Published online** 22, September, 2011

**Published in print edition** September, 2011

This book on power quality written by experts from industries and academics from various countries will be of great benefit to professionals, engineers and researchers. This book covers various aspects of power quality monitoring, analysis and power quality enhancement in transmission and distribution systems. Some of the key features of books are as follows: Wavelet and PCA to Power Quality Disturbance Classification applying a RBF Network; Power Quality Monitoring in a System with Distributed and Renewable Energy Sources; Signal Processing Application of Power Quality Monitoring; Pre-processing Tools and Intelligent Techniques for Power Quality Analysis; Single-Point Methods for Location of Distortion, Unbalance, Voltage Fluctuation and Dips Sources in a Power System; S-transform Based Novel Indices for Power Quality Disturbances; Load Balancing in a Three-Phase Network by Reactive Power Compensation; Compensation of Reactive Power and Sag Voltage using Superconducting Magnetic Energy Storage; Optimal Location and Control of Flexible Three Phase Shunt FACTS to Enhance Power Quality in Unbalanced Electrical Network; Performance of Modification of a Three Phase Dynamic Voltage Restorer (DVR) for Voltage Quality Improvement in Distribution System; Voltage Sag Mitigation by Network Reconfiguration; Intelligent Techniques for Power Quality Enhancement in Distribution Systems.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Zahra Moravej, Mohammad Pazoki and Ali Akbar Abdoos (2011). Application of Signal Processing in Power Quality Monitoring, Power Quality – Monitoring, Analysis and Enhancement, Dr. Ahmed Zobaa (Ed.), ISBN: 978-953-307-330-9, InTech, Available from: <http://www.intechopen.com/books/power-quality-monitoring-analysis-and-enhancement/application-of-signal-processing-in-power-quality-monitoring>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821



© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen